PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Oral Pre-exposure Prophylaxis (PrEP) to prevent HIV: a systematic review and meta-analysis of clinical effectiveness, safety, adherence and risk compensation in all populations
AUTHORS	O Murchu, Eamon; Marshall, Liam; Teljeur, Conor; Harrington, Patricia; Hayes, Catherine; Moran, Patrick; Ryan, Mairin

VERSION 1 – REVIEW

REVIEWER	Larmarange, Joseph Centre Population et Développement, Institut de Recherche pour le Développement, Université Paris Descartes, ERL Inserm U 1244
REVIEW RETURNED	22-Feb-2021

GENERAL COMMENTS	This paper presents a systematic review and meta-analysis about the effectiveness of oral pre-exposure prophylaxis to prevent HIV acquisition. The authors also reviewed safety, adherence, and risk compensation. The analysis is clearly and well performed. The authors followed PRISMA guidelines, and results are reported according to the GRADE framework. However, to avoid any misinterpretation by non-technical readers and for clarity, I feel that the paper could benefit from minor revisions, particularly considering how the results could be interpreted in terms of public health implications. It is also essential to better show what is known but also what is not (yet) known. For example, it should be more explicit in the abstract (and in the manuscript) that the effectiveness of PrEP is measured according to intent-to-treat analysis and not per-protocol and therefore depends on both the efficacy of the drug itself and on the level of adherence. Considering the importance of adherence on PrEP effectiveness, as shown by the authors in the result, it would be helpful to add two columns in Table 1 to indicate which are the studies where adherence was high (>80%) and which are the studies where PrEP was effective. For example, it appears that among the five studies classified as “heterosexual transmission”, only one was classified “high adherence” (Thigpen 2012), and this is the only one to show any effectiveness of PrEP. As highlighted by the authors, the study of Thigpen 2012 is also specific as it included both men and women, and when looking at results per gender, PrEP was effective (ITT analysis) only among men. However, when looking at the “as-treated” analysis (table S8b of the supplementary materials of Thigpen 20212), PrEP

	appears effective among women, indicating that adherence was not similar between gender. Considering that it could have different program implications, it would be relevant to stratify Table 3 by adherence level. In its current form, the authors stratified efficacy results by adherence only for MSM (but showing only the high adherence subgroup). Showing, for each population, high adherence and low adherence subgroup results would provide a much more global and comprehensive overview. Some rows would be empty, showing what is currently undocumented/unknown. Regarding the subpopulations considered by the authors, some categories should be clarified. It seems that the initial focus of the analysis was MSM before being extended to other populations (in supplementary material S1, section 3.1: line 14 restricts the analysis to MSM before extending to all populations line 19). “Serodiscordant couples” are in fact “serodiscordant couples when the positive partner is not on ART”. No study showed a benefit of PrEP when the positive partner is on ART. Such benefit is also not expected due to the fact the preventive effect of ART. In terms of recommendations, it is therefore unclear if PrEP should still be recommended once the HIV-infected partner has achieved viral suppression. Regarding “heterosexual transmission”, it would be better to distinct men and women. It should also be mentioned “in high-prevalence context”: all reported studies were conducted in Southern Africa, and it is unclear if PrEP would be beneficial in the general population in low-prevalence settings. Considering specific risks and vulnerabilities faced by female sex workers (FSW), it is surprising that the authors did not consider such sub-population. According to Table 2 and if I understood correctly, the authors did not find any study specifically among FSW, and only one study (Peterson 2007) included FSW but did not report results per subgroups. The absence of data for that specific population should be highlighted in the Discussion and the conclusion. Additional clarification regarding methods: according to Figure 1, 2803 records have been identified. 2730 were excluded. Could the reasons for exclusion be clarified?

REVIEWER	White, Ellen University College London, Institute of Clinical Trials and Methodology
REVIEW RETURNED	26-Apr-2021

GENERAL COMMENTS	Abstract • Please report the estimated effectiveness (and confidence intervals) of PrEP in heterosexuals, rather than stating that the effect is non-significant. • As described below, please clarify what is meant by >=80% adherence in a given study. • The authors define RR as relative risk in the abstract, but this is inconsistent with the rest of the manuscript. Introduction • The introduction is missing a number of important references, such as reference to the PROUD and IPERGAY study when discussing the PrEP studies in MSM and the studies in heterosexual women in the following sentence (e.g. FEMPrEP trial).

	• The last sentence of the introduction describes that this review aimed to inform the Irish government. However, the article summary suggests that study has assisted development of clinical practice – please clarify. Methods • Please could the authors expand upon what they mean by hand-searching journals? • Were studies in serodiscordant couples solely heterosexual couples? If so, please specify in methods. • Please clarify which analyses were stratified according to the four subgroups. It is incorrect to say all analyses were stratified by population, as a number of analyses have been presented across all populations, e.g. safety analyses. • The authors report that risk ratios were used, despite describing the use of person-years. Please correct throughout the manuscript if in fact rate ratios are used and ensure that methods are appropriate for this. • The authors report that the modified intention to treat analysis removed the number of participants from the denominator. However, this should exclude the person-years from the denominator and number with a HIV infection from the numerator. Please clarify what approach was used. • The authors describe that a sensitivity analysis was conducted for analyses where no events had occurred. Please could you expand on what this analysis was aiming to ascertain? • High adherence was described as those studies with >=80% of participants adhering to study medication, but it is unclear the definition used to define which participants were adherent. Please make clearer throughout manuscript. Results • Table 2 – PROUD contained 544 participants – please correct and check others. • Table 3 – Please check the symbols in GRADE column for heterosexual studies. This is labelled as LOW evidence, but three symbols are crossed. • Forest plots - please ensure that labelling is clear, i.e. “Event” are HIV diagnoses and “Total” is person-year(?). • Presenting the overall effectiveness on page 17 is unnecessary given that the authors specified a priori that there was likely to be heterogeneity between the subgroups. • The sentence on adherence by plasma drug detection is an unusual place (page 17) given that the adherence is described further down in the results section. Please consider moving. • When discussing the sensitivity analyses for effectiveness, it is unclear whether these results are presented elsewhere. If so, please refer to table/figure. If not, please provide effect estimate of RR and confidence intervals. • Please clarify for Figure S4 how adherence was defined for using drug plasma concentrations. This is particularly important when interpreting what we mean by adherence throughout this manuscript. PROUD, for example, only performed drug concentrations in those that reported they were taking PrEP and found that all had drug detectable. I would therefore argue that studies with this kind of selective bias should not be included within the adherence analyses. A bigger issue, however, is that it is not clear what figure 4 is reporting given that maximum adherence is ~88% (which was defined according to prescriptions in PROUD, for example), especially since the author reports that pill counts were not used. Please clarify/correct.

	• I think the first sentence in the “Relationship between efficacy and adherence” section is unnecessary given that the authors subsequently report a meta-regression. • Please reference studies when referring to them in results. Discussion • The summary of findings could be much more succinct to get the key messages across. • It is important that these findings are put into context of the population at risk of HIV in Ireland. • Please reference the guidance/policy that the study has contributed to. It’s also important to know how these results informed the Irish government. • Pg. 28 – please check whether this sentence needs revising “Second, while risk compensation was not noted in this review,….”

REVIEWER	Dronavalli, Mithilesh University of Western Australia, WACRH
REVIEW RETURNED	16-Jun-2021

GENERAL COMMENTS	Excellent research question which is highly applicable to public health policy and HIV prevention. My suggestions are: Adherence should be as a continuous variable rather than a dichotomy at 80% and 20%. I recommend doing meta-regression to determine the effectiveness of PrEP adjusting for adherence as a continuous variable in the cohorts of MSM and Sero-discordant and all studies. Modified intention to treat is not as good as intention to treat. State the type of modified intention to treat utilised for each study. I recommend a sensitivity analysis for on demand versus daily PrEP. Daily and on-demand RCT arms may be quite different in effectiveness. This would add significantly to the paper, without the need to collect additional data. Finally, the study by Bekker in 2018 on the Heterosexual group of people should be excluded as it compares two types of PreP, which is contrary to the objective of the study: "To conduct a systematic review and meta-analysis of randomised controlled trials (RCTs) on the effectiveness and safety of Pre-Exposure Prophylaxis (PrEP) to prevent HIV." This is also why I have recommended doing a sensitivity analysis for studies including and excluding intermittent/on-demand dosing. Good Luck, great cause!

REVIEWER	Dronavalli, Mithilesh University of Western Australia, WACRH
REVIEW RETURNED	25-Jun-2021

GENERAL COMMENTS	Major points: - Please add the summary of search strategy and main keywords in the abstract. - The search step needs to be updated (July 2020 !!). - Please mention the results of Heterogeneity (Q or chi-square test) and publication bias tests in the main text for each of studied outcome.

	Minor points: - Please mention Kappa agreement coefficient between two independent reviewers in the method section - The results of sensitivity analysis should be reported in the results section. - Please add funnel plots for assessing the symmetric assumption for each of studied outcome - Please check the last column of Table S3.4. It seem that this column is p-value.

REVIEWER	Li, Chunyan University of North Carolina at Chapel Hill, Health Behavior
REVIEW RETURNED	15-Aug-2021

GENERAL COMMENTS	1. Categorization of populations: the current four categories listed in the manuscript include MSM, PWID, heterosexuals, and serodiscordant couples. In the main text, this is no mention of how the authors categorized transgender populations. But in Table 2, I found the authors defined transgender women as MSM, and transgender men as heterosexuals. While I don't think transgender people should be merged into either of the groups, could the authors please explain their rationale of doing so? Is it possible (it may not) to separate data of transgender people from the rest and analyze the outcome measures? 2. It surprised me when reading that the authors conclude that PrEP is only effective for MSM, while WHO and many countries have recommended PrEP for all populations at substantial risks of HIV infection for many years. The 2015 WHO guideline on PrEP has a great section of efficacy evidence synthesis that listed all the trial data and evidence of PrEP efficacy in key populations, including women. In the context of adherence-based analysis, evidence is clear that oral PrEP is effective in reducing chances of HIV infection among high-adherent individuals, regardless of group. A Lancet article that pooled data of transgender women from two clinical trials also showed the efficacy of PrEP in HIV prevention. (DOI: https://doi.org/10.1016/S2352-3018(15)00206-4) 3. PrEP has rapidly evolved during the past years, including the development of formulations and delivery modalities. It looks like the authors focused on oral regimens only. I would make this point more clear by indicating/stating early in the title and abstract. 4. Some statements in the Introduction may need to be updated. For example, the latest estimated number of new HIV infections globally is 1.5 million in 2020, which could be found on UNAIDS's website.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:			
1	This paper presents a systematic review and meta-analysis about the effectiveness of oral pre-exposure prophylaxis to prevent HIV acquisition. The authors also reviewed safety, adherence, and risk compensation. The analysis is clearly and well performed. The authors followed PRISMA guidelines, and results are reported according to the GRADE framework. However, to avoid any misinterpretation by non-technical readers and for clarity, I feel that the paper could benefit from minor revisions, particularly considering how the results could be interpreted in terms of public health implications. It is also essential to better show what is known but also what is not (yet) known. For example, it should be more explicit in the abstract (and in the manuscript) that the effectiveness of PrEP is measured according to intent-to-treat analysis and not per-protocol and therefore depends on both the efficacy of the drug itself and on	We thank the reviewer for these comments. In the abstract, we have added that the analysis was a modified intention-to-treat analysis (all HIV negative participants at enrolment were included in analyses). Additionally, in the methods and results sections, we have further described the analysis type and highlighted that in the modified intention-to-treat analysis, as opposed to a per-protocol analysis, the effectiveness was a function of both efficacy of the drug itself and on adherence.	• Abstract, Page 2 • Methods section, Page 9, 2nd paragraph, under sub-heading 'Data collection and analysis' • Results section, Page 13

	the level of adherence.		
2	Considering the importance of adherence on PrEP effectiveness, as shown by the authors in the result, it would be helpful to add two columns in Table 1 to indicate which are the studies where adherence was high (>80%) and which are the studies where PrEP was effective. For example, it appears that among the five studies classified as "heterosexual transmission", only one was classified "high adherence" (Thigpen 2012), and this is the only one to show any effectiveness of PrEP. Considering the importance of adherence on PrEP effectiveness, as shown by the authors in the result, it would be helpful to add two columns in Table 1 to indicate which are the studies where adherence was high (>80%) and which are the studies where PrEP was effective. For example, it appears that among the five studies classified as "heterosexual transmission", only one was classified "high adherence" (Thigpen 2012), and this is the only one to	We assume the reviewer is referring to Table 2 (as Table 1 is the PICOS table of study inclusion criteria in the methods section). Table 2 is the "table of study characteristics", which precedes our sections on effectiveness and safety. We have included an additional column in Table 2 that reports study adherence, per the reviewer's comments. However, we would not consider it standard practice to report effectiveness/safety outcomes in this table, as these are presented In the following table (Table 3 – GRADE summary of findings tables for effectiveness and safety).	Table 2, Pages 14-17

	show any effectiveness of PrEP.		
3	Considering that it could have different program implications, it would be relevant to stratify Table 3 by adherence level. In its current form, the authors stratified efficacy results by adherence only for MSM (but showing only the high adherence subgroup). Showing, for each population, high adherence and low adherence subgroup results would provide a much more global and comprehensive overview. Some rows would be empty, showing what is currently undocumented/unknown.	We have amended Table 3 to include all adherence groups. Therefore, we have added MSM 'low adherence', heterosexual transmission 'high adherence', and heterosexual transmission 'low adherence'. Therefore, MSM and heterosexual populations each have three subgroups: all studies, high adherence and low adherence. The remaining risk groups relate to serodiscordant couples (2 studies with high adherence) and PWID (1 study with low adherence). We have amended the text of the table to indicate the adherence level within these groups. We feel that an empty row for 'low adherence' is not necessary for serodiscordant couples, as the goal of PrEP trials is to achieve high adherence. Additionally, for PWID, despite low adherence PrEP was found to be effective; an empty row for 'high adherence' is unlikely to add to the table in a meaningful way. Therefore, in the revised table, we have rows for 'all studies' for each population group, and when there is a mix of trials with both low and high adherence, we now have subgroups that report the GRADE findings separately. For serodiscordant couples and PWID, PrEP was found to be effective, and studies within these categories all belonged to the same adherence subgroup and therefore additional rows were not deemed necessary. Additional text has been added to add clarity relating to adherence in these populations.	Table 3, Pages 19,20
4	Regarding the subpopulations considered by the authors, some categories should be clarified. It seems that the initial focus of the	This was an error in the protocol that was submitted as part of the Supplementary Material – we have amended this sentence. Initially, the study was only focussed on MSM, however following discussion with the stakeholder (the Irish	Supplementary Material 3

	analysis was MSM before being extended to other populations (in supplementary material S1, section 3.1: line 14 restricts the analysis to MSM before extending to all populations line 19).	Department of Health), the focus of the study was expanded to include other at-risk groups.	
5	"Serodiscordant couples" are in fact "serodiscordant couples when the positive partner is not on ART". No study showed a benefit of PrEP when the positive partner is on ART. Such benefit is also not expected due to the fact the preventive effect of ART. In terms of recommendations, it is therefore unclear if PrEP should still be recommended once the HIV-infected partner has achieved viral suppression.	We agree with these comments, and Irish guidelines on HIV prevention do not recommend PrEP in serodiscordant couples when the HIV-positive partner is stably suppressed on antiretroviral medications. This has been clarified in the report (studies on serodiscordant couples only relate to couples when the partner is not on ART).	• Methods section, Page 7, 3rd paragraph • Results section, Page 23, under subheading 'Effectiveness in serodiscordant couples' • Tables 2 and 3
6	Regarding "heterosexual transmission", it would be better to distinct men and women. It should also be mentioned "in high-prevalence context": all reported studies were conducted in Southern Africa, and it is unclear if PrEP would be beneficial in the general population in low-prevalence settings.	We have amended the reporting of heterosexual transmission to include these points. We have clarified that four of the five studies enrolled only women, and one study enrolled both men and women. The study that enrolled men and women was the only study with high adherence. In this study, effectiveness was only found among men (we report disaggregated data for this study). Additionally, we have highlighted that all studies were conducted in a high HIV prevalence context (countries in Sub-Saharan Africa).	• Methods section, Page 7, last paragraph • Results section, Pages 23,24, under subheading 'Effectiveness in heterosexuals') • Tables 2 and 3
7	Considering specific risks and vulnerabilities faced by female sex workers (FSW), it is surprising that the authors did	Unfortunately, we did not identify any study that reported efficacy specifically among female sex workers (the study by Peterson et al. included female sex workers, but did not report disaggregated data). This	• Abstract • Discussion section, Page 32, 2nd paragraph, under subheading 'Strengths and limitations'

	not consider such sub-population. According to Table 2 and if I understood correctly, the authors did not find any study specifically among FSW, and only one study (Peterson 2007) included FSW but did not report results per subgroups. The absence of data for that specific population should be highlighted in the Discussion and the conclusion	limitation of the review has been emphasised in the discussion and has been added as a limitation of the review in the abstract.	
8	Additional clarification regarding methods: according to Figure 1, 2803 records have been identified. 2730 were excluded. Could the reasons for exclusion be clarified?	We have added a footnote relating to the reason for excluding these studies (at the stage of screening title/abstract, 2730 citations were excluded as they did not meet our inclusion/exclusion criteria, per our PICOS). This step preceded full text review of potentially relevant articles. In our supplementary material, we have listed each study excluded based on full text review along with the reason for exclusion.	Footnote to Figure 1
Reviewer #2:			
1	Abstract Please report the estimated effectiveness (and confidence intervals) of PrEP in heterosexuals, rather than stating that the effect is non-significant.	We have added the estimate of effectiveness in heterosexuals.	Abstract, Page 2
	As described below, please clarify what is meant by >=80% adherence in a given study.	We have added greater detail on this variable in the methods section (how adherence was estimated and how studies were categorised as 'high' or 'low' adherence). Subgroups of adherence groups have now been removed from the abstract, as it was necessary to reduce the text in the abstract to	Methods section, Pages 9,10, under sub-heading 'Data collection and analysis'

		comply with the word count (as there have been additions in other places).	
	The authors define RR as relative risk in the abstract, but this is inconsistent with the rest of the manuscript.	Our estimate of effect was the rate ratio, which can be interpreted in the same way as relative risk or risk ratio. We have changed 'relative risk/risk ratio' to 'rate ratio' throughout.	Throughout
2	Introduction The introduction is missing a number of important references, such as reference to the PROUD and IPERGAY study when discussing the PrEP studies in MSM and the studies in heterosexual women in the following sentence (e.g. FEMPrEP trial).	We have included reference to PROUD, IPERGAY and FEM-PrEP in the introduction, per the reviewer's comments. We have only included a minimum number of key references in the introduction, however, as studies on PrEP efficacy/safety are the output of the review that will be identified by the database search.	Introduction, Page 5, 2nd paragraph
3	The last sentence of the introduction describes that this review aimed to inform the Irish government. However, the article summary suggests that study has assisted development of clinical practice – please clarify.	The last sentence of the introduction is as follows: This review aimed to inform the decision of the Irish government to implement a PrEP programme and to assist in the development of national clinical practice guidelines on PrEP for HIV prevention. The purpose of the review was twofold: to assist the Irish Department of Health in its decision whether or not to implement a national Pre-Exposure Prophylaxis programme, and to assist the Health Service Executive and the Sexual Health and Crisis Pregnancy Programme to develop clinical practice guidelines relating to HIV prevention. The references for the Irish government's decision to implement a PREP programme, and reference to the clinical practice guidelines, have been added for clarity in the discussion section.	Discussion section, Page 32, 1st paragraph, References 31 and 32
4	Methods Please could the authors expand upon what they mean by	This sentence was unnecessary and has been removed. It relates to our standard research methods that include the hand-searching hard	

	hand-searching journals?	copies of journals and conference proceedings if they are unavailable online. This was not necessary in this review.	
5	Were studies in serodiscordant couples solely heterosexual couples? If so, please specify in methods.	Yes – studies in serodiscordant couples only enrolled heterosexual serodiscordant couples. This has been clarified in the methods and results section.	• Methods section, Page 7, last paragraph • Results section, Pages 23,24, under subheading 'Effectiveness in heterosexuals' • Tables 2 and 3
6	Please clarify which analyses were stratified according to the four subgroups. It is incorrect to say all analyses were stratified by population, as a number of analyses have been presented across all populations, e.g. safety analyses.	This has been amended. All effectiveness analysis were stratified by population. Safety analyses were not stratified as it was not considered likely that the safety profile would differ between populations. In addition, presenting results across all populations gave greater power to detect rare safety events.	Methods section, Page 7, last paragraph
7	The authors report that risk ratios were used, despite describing the use of person-years. Please correct throughout the manuscript if in fact rate ratios are used and ensure that methods are appropriate for this.	The terminology has been amended. The rate ratio is the appropriate terminology (which is the rate of HIV infection in the PrEP group compared with control). The rate of HIV infection represented the number of HIV infections that occurred per person-years of follow up data, and the RR compares the rate of HIV infection in the PrEP group with control. The rate of HIV infection (per person-years) was favoured over risk of HIV infection as rate incorporates both the number of participants and the duration of follow-up, allowing for comparisons across studies that may vary significantly in terms of study duration. The interpretation of the rate ratio is similar to the risk ratio. The methods employed are appropriate for this.	Throughout
8	The authors report that the modified intention to treat analysis removed the number of participants from the denominator. However, this should	The explanation of the modified intention-to-treat analysis has been amended to add clarity. A modified intention-to-treat analysis was selected instead of a standard intention-to-treat analysis to account for unrecognised HIV infection at	Methods section, Page 9, 2nd paragraph, under sub-heading 'Data collection and analysis'

	exclude the person-years from the denominator and number with a HIV infection from the numerator. Please clarify what approach was used.	enrolment. Therefore, all patients who were HIV negative at enrolment in the study were included in analyses. The numerator was the number of new HIV infections (after enrolment) and the denominator was the number of person-years at risk in HIV negative participants at enrolment. Only patients with an unrecognised HIV infection at enrolment were excluded from the numerator and denominator.	
9	The authors describe that a sensitivity analysis was conducted for analyses where no events had occurred. Please could you expand on what this analysis was aiming to ascertain?	We have expanded on this concept in the methods section. Trials with zero events in both arms are typically excluded in meta-analyses, resulting in a loss of information. Approaches are available to include zero event trials with application of a continuity correction (whereby all cells in the two by two table for a given study have 0.5 added to avoid division by zero). This approach can also lead to bias, particularly for small trials or those with imbalanced arms. For this study, if trials with zero events in one or both arms were identified, a sensitivity analysis using a random effects Poisson regression and beta-binomial models was applied to determine whether the results were sensitive to presence of trials with zero events in one or both arms. The main analysis, however, followed the standard approach of excluding trials with zero events in both arms.	Methods section, Page 11, 1st paragraph, under sub-heading 'Data collection and analysis'
10	High adherence was described as those studies with >=80% of participants adhering to study medication, but it is unclear the definition used to define which participants were adherent. Please make clearer throughout manuscript	The manuscript has been amended to clarify our definition of adherence. For categorising studies as 'high' or 'low' adherence, we were limited by the methods of determining adherence used in the primary studies which varied considerably across studies. Adherence was measured in a number of ways, including self-report, pill counts, medication event monitoring systems (MEMS), structured interviews and plasma drug detection methods. In studies that used a variety of	• Methods section, Page 9, last paragraph and Page 10, 1st paragraph, under sub-heading 'Data collection and analysis' • Results section, Page 25, under sub-heading 'Relationship between efficacy and adherence'

		methods, we favoured plasma drug concentrations over pill-counts or self report, as plasma drug concentration was considered the least biased measurement. In studies that reported multiple measures of adherence without reporting plasma drug concentration, we used the lowest reported adherence (taking a conservative approach, as adherence was frequently over-reported in trials). Only trials that reported plasma drug detection were included in the meta-regression analysis.	
11	Results Table 2 – PROUD contained 544 participants – please correct and check others.	This has been amended.	Table 2, Page 14
12	Table 3 – Please check the symbols in GRADE column for heterosexual studies. This is labelled as LOW evidence, but three symbols are crossed.	This has been amended.	Table 3, Page 19
13	Forest plots - please ensure that labelling is clear, i.e. “Event” are HIV diagnoses and “Total” is person-year(?).	This is correct. We have amended the forest plots to include this description in the footnote – ‘events’ refers to new HIV infections after enrolment in the study, and ‘total’ refers to person-years at risk during the study period.	Figure 2: Forest plot – footnote
14	Presenting the overall effectiveness on page 17 is unnecessary given that the authors specified a priori that there was likely to be heterogeneity between the subgroups.	This has been amended. While we found a forest plot of all trials useful for visual inspection of individual studies, we agree that there is great heterogeneity and the summary estimate is not very meaningful. As we are now removing the section on overall effectiveness, the forest plot with ‘all trials’ has also been removed, which has changed the numbering of subsequent figures.	Results section, Page 18 (under subheading ‘Effectiveness’)
15	The sentence on adherence by plasma drug detection is an	This sentence has been removed from this position.	Results section, Page 18 (under subheading ‘Effectiveness’)

	unusual place (page 17) given that the adherence is described further down in the results section. Please consider moving.		
16	When discussing the sensitivity analyses for effectiveness, it is unclear whether these results are presented elsewhere. If so, please refer to table/figure. If not, please provide effect estimate of RR and confidence intervals.	We have expanded our reporting of the sensitivity analysis, and it is now reported in a separate section in the Results section. Location of changes: Results section	Results section, Pages 24,25 (under subheading 'Sensitivity analysis')
17	Please clarify for Figure S4 how adherence was defined for using drug plasma concentrations. This is particularly important when interpreting what we mean by adherence throughout this manuscript. PROUD, for example, only performed drug concentrations in those that reported they were taking PrEP and found that all had drug detectable. I would therefore argue that studies with this kind of selective bias should not be included within the adherence analyses. A bigger issue, however, is that it is not clear what figure 4 is reporting given that maximum adherence is ~88% (which was defined according to prescriptions in PROUD, for example), especially	In the meta-regression analysis, we excluded studies that did not undertake plasma drug detection to confirm adherence. This has been clarified in the report ("Studies that did not confirm adherence through plasma drug detection rates were excluded from analyses"). In the PROUD study, plasma drug detection was only carried out on participants who reported taking the study drug (88% of participants reported taking study drug). 100% of these samples had study drug detected. We considered 88% the correct estimate of adherence; in this case, plasma drug detection confirmed adherence among the 88% who reported taking study drug. This method of calculating adherence has been clarified in the methods section ("In studies that only measured plasma drug concentration in participants who reported taking study drug, confirmed adherence was calculated by multiplying the proportion of samples with study drug detected by the self-reported adherence rate"). Additionally, the method of determining adherence in PROUD study has been added to Table 2.	• Methods section, Page 10, 1st paragraph (under sub-heading 'Data collection and analysis') • Results section, Page 25 (under sub-heading 'Relationship between efficacy and adherence')

	since the author reports that pill counts were not used. Please clarify/correct.		
18	I think the first sentence in the "Relationship between efficacy and adherence" section is unnecessary given that the authors subsequently report a meta-regression.	This sentence has been removed.	Results section, Page 25 (under sub-heading 'Relationship between efficacy and adherence')
19	Please reference studies when referring to them in results.	The results section has been revised to ensure references have been inserted throughout.	Results section (throughout)
20	Discussion The summary of findings could be much more succinct to get the key messages across	Any unnecessary text in the summary of findings has been removed.	Discussion section, Page 30 (under sub-heading 'Summary of findings')
21	It is important that these findings are put into context of the population at risk of HIV in Ireland.	An additional paragraph has been added to the discussion that outlines the context of HIV transmission in Ireland. Note that 2018 data were used as these are the most recent published data available.	Discussion section, Pages 33,34 (under sub-heading 'Research in context and implications for practice')
22	Please reference the guidance/policy that the study has contributed to. It's also important to know how these results informed the Irish government	The reference to the national clinical guidelines have been inserted, in addition to a reference relating to the Irish government's announcement that a publicly funded PrEP programme would be implemented. The study was used to develop national clinical guidelines on the management of patients on PrEP, and it informed the decision of the Irish government to implement a publicly funded PrEP programme nationally for MSM and serodiscordant couples at increased risk, and for other populations on a case-by-case basis as determined by the treating HIV specialist.	Discussion section, Page 32, 1st paragraph, References 31 and 32 (under sub-heading 'Strengths and limitations')
23	Pg. 28 – please check whether this sentence needs revising "Second, while risk	This sentence has been amended to add clarity.	Discussion section, Page 32, last paragraph, 1st sentence (under sub-heading 'Strengths and limitations')

	compensation was not noted in this review,…."		
Reviewer #3:			
1	Adherence should be as a continuous variable rather than a dichotomy at 80% and 20%. I recommend doing meta-regression to determine the effectiveness of PrEP adjusting for adherence as a continuous variable in the cohorts of MSM and Sero-discordant and all studies.	It may have been overlooked by the reviewer that adherence was a continuous variable in the meta-regression analysis. This has now been made explicit in the methods section. Separately, in our subgroup analysis we defined high and low risk groups as a categorical variable (≥80 and <80% adherence based on proportion of participants adherent to study drug) for each population.	Methods section, Page 10, 2nd paragraph (under sub-heading 'Data collection and analysis')
2	Modified intention to treat is not as good as intention to treat. State the type of modified intention to treat utilised for each study.	All primary studies reported the modified intention-to-treat analysis, which was selected over per-protocol analysis. The definition of modified intention-to-treat analysis was the same in each study: all patients who were HIV negative at enrolment in the study were included, and patients who were HIV positive at baseline due to an unrecognised HIV infection at enrolment were excluded. We have clarified our modified intention-to-treat analysis in the text. In this review, we did not consider a standard intention-to-treat approach appropriate as it did not account for participants who were already HIV positive prior to enrolment.	Methods section, Page 9, 2nd paragraph (under sub-heading 'Data collection and analysis')
3	I recommend a sensitivity analysis for on demand versus daily PrEP. Daily and on-demand RCT arms may be quite different in effectiveness. This would add significantly to the paper, without the need to collect additional data.	Only one study administered 'on demand' PrEP (IPERGAY trial) using the dosing regimen advocated by international clinical guidelines (https://www.cdc.gov/hiv/basics/prep/on-demand-prep.html). We have updated the results and discussion sections to describe this study separately (IPERGAY study). In this study with high adherence conducted among MSM, the efficacy was 86%, identical to the most comparable high adherence study	• Results section, Page 22, 2nd paragraph (under sub-heading 'Effectiveness in MSM') • Discussion section, Page 30, 1st paragraph (under sub-heading 'Summary of findings')

		among MSM using daily dosing (PROUD study, also 86%). Two studies used 'intermittent' dosing, however the dosing regimens varied significantly and pooling data was not considered appropriate (and we did not consider these studies 'on demand' PrEP). In addition, the data would not contribute to a sensitivity analysis in a meaningful way. The studies are: • Mutua et al.: Daily/intermittent PrEP was compared with daily /intermittent placebo in MSM. The intermittent dosing schedule in this study was Monday and Friday, and 2 hours after sex. This dosing schedule was too dissimilar to the IPERGAY trial to pool data. In this small study, there were no infections in the daily or intermittent groups (and one case in the placebo group). • Kibengo et al.: Daily/intermittent PrEP was compared with daily/intermittent placebo in serodiscordant couples. However, no events (HIV infections) occurred in either arm and therefore RR could not be estimated.	
4	Finally, the study by Bekker in 2018 on the Heterosexual group of people should be excluded as it compares two types of PreP, which is contrary to the objective of the study: "To conduct a systematic review and meta-analysis of randomised controlled trials (RCTs) on the effectiveness and	We have clarified that the objective was to investigate the effectiveness and safety of oral tenofovir-containing PrEP compared with placebo, no treatment or alternative oral PrEP medication/dosing schedule. Our protocol included studies that compare different types of PrEP or alternative dosing schedules, as these studies can provide safety and relative effectiveness data (and the comparator is contained in our PICOS criteria – Table 1). No study directly compared 'on demand' PrEP	Abstract, Page 2

	safety of Pre-Exposure Prophylaxis (PrEP) to prevent HIV." This is also why I have recommended doing a sensitivity analysis for studies including and excluding intermittent/on-demand dosing.	(using the definition of 'on demand' in IPERGAY trial) with daily PrEP, which would have been a very useful comparison. While we have evidence from the IPERGAY study that on demand PrEP is highly effective (86%), we have no direct evidence to inform a relative effectiveness assessment.	
Reviewer #4:			
1	Major points: Please add the summary of search strategy and main keywords in the abstract.	The list of the databases searched as part of our search strategy has been added to the abstract, and the main keywords for the review have been included at the end of the abstract. Our description of the search strategy has been expanded in the methods section – we now include the main search terms of the database search.	• Abstract, Pages 2,3 • Methods section, Page 7, 2nd paragraph (under sub-heading 'Search strategy and selection criteria')
2	The search step needs to be updated (July 2020 !!)	We have performed a literature review to identify recently published or ongoing studies since the date of our original database search. This search has been added to the discussion under the subheading 'ongoing studies'. For this search, PubMed was searched up to 9 September 2021 using the same search strategy as the original search. No publications were identified that assessed the effectiveness or safety of PrEP compared with placebo or no treatment. Two publications relating to one ongoing study was identified that compared two different salts of tenofovir, however (tenofovir alafenamide vs. tenofovir disoproxil fumarate). Tenofovir alafenamide was found to be non-inferior to tenofovir disoproxil fumarate in the interim analyses of this ongoing trial.	Discussion section, Page 31 (under sub-heading 'Ongoing studies')
3	Please mention the results of Heterogeneity (Q or	The I^2 value was the measure of heterogeneity used in analyses, per the study protocol. The I^2 value has	The I^2 value has been reported next to all outputs from meta-analyses in the Results section. The forest plot

	chi-square test) and publication bias tests in the main text for each of studied outcome.	been included for all reported meta-analyses. The forest plot also includes the results of the Chi-squared test. The I^2 value was used to evaluate heterogeneity in our GRADE assessment. We do not consider publication bias tests to be appropriate for each outcome, as each outcome was reported in fewer than 10 studies (the largest number of studies was in the MSM population: n=6 studies). See comment number 6, below, relating to the suitability of funnel plots and statistical tests for publication bias.	already includes the I^2 value and Chi-square test.
4	Minor points: Please mention Kappa agreement coefficient between two independent reviewers in the method section	We did not measure Kappa agreement in this review, per best practice guidelines. We followed the methodology for the conduct of systematic review outlined in the Cochrane handbook (Chapter 7, Section 7.3.2 Performing assessments of risk of bias, https://training.cochrane.org/handbook/current/chapter-07): "We do not recommend the use of statistical measures of agreement (such as kappa statistics) to describe the extent to which assessments by multiple authors were the same. It is more important that reasons for any disagreement are explored and resolved."	
5	The results of sensitivity analysis should be reported in the results section.	We have expanded our reporting of the sensitivity analysis, and it is now reported in a separate section in the Results section.	Results section, Pages 24,25, under subheading 'Sensitivity analysis'
6	Please add funnel plots for assessing the symmetric assumption for each of studied outcome	For each outcome, we do not consider funnel plots to be appropriate, as each outcome had too few studies (the maximum number of studies was 6 in the MSM group). As mentioned previously, our systematic review followed Cochrane methodology. Advice on the use of tests for funnel plot asymmetry is provided in Chapter 10 of the Cochrane handbook (https://handbook-5-	• Results section, Page 18, 1st paragraph • Supplementary Material 2.3

		1.cochrane.org/chapter_10/10_4_3_1_recommendations_on_testing_for_funnel_plot_asymmetry.htm): "For all types of outcome: As a rule of thumb, tests for funnel plot asymmetry should be used only when there are at least 10 studies included in the meta-analysis, because when there are fewer studies the power of the tests is too low to distinguish chance from real asymmetry." However, a funnel plot for 'all studies' may be of relevance (n=13 studies). A limitation of this approach is that ideally, publication bias should be assessed for each reported outcome. We have added this funnel plot in Supplementary Material 2.3. There was no asymmetry noted. Additionally, the arcsine test for funnel plot asymmetry was applied to all 13 trials. The p-values for the equivalent of the Begg, Egger and Thompson tests were 0.58, 0.14 and 0.13, respectively. This has been added to the results section.	
7	Please check the last column of Table S3.4. It seem that this column is p-value.	This has been amended – the last column is the p-value.	Table in Supplementary Material 2.5
Reviewer #5:			
1	Categorization of populations: the current four categories listed in the manuscript include MSM, PWID, heterosexuals, and serodiscordant couples. In the main text, this is no mention of how the authors categorized transgender populations. But in Table 2, I found the authors defined transgender women as MSM, and	Unfortunately, no data were identified relating specifically to transgender women or men. In Table 2, there were two trials in which transgender individuals were eligible for enrolment. The categorisation is as follows: • iPrEX trial – categorised as MSM iPrEX was categorised as MSM due to the very small number of transgender women: 100% were male at birth with 1% female gender identity. Data were not disaggregated to provide an analysis of transgender women.	Table 2, Page 16, Bekker 2018 trial

	transgender men as heterosexuals. While I don't think transgender people should be merged into either of the groups, could the authors please explain their rationale of doing so? Is it possible (it may not) to separate data of transgender people from the rest and analyze the outcome measures?	• Bekker 2018 – categorised as heterosexual (women) While transgender men were eligible for this study, all participants were in fact cis-gendered. Therefore this study only relates to heterosexual women. Table 2 has been amended to remove 'Transgender men'.	
2	It surprised me when reading that the authors conclude that PrEP is only effective for MSM, while WHO and many countries have recommended PrEP for all populations at substantial risks of HIV infection for many years. The 2015 WHO guideline on PrEP has a great section of efficacy evidence synthesis that listed all the trial data and evidence of PrEP efficacy in key populations, including women. In the context of adherence-based analysis, evidence is clear that oral PrEP is effective in reducing chances of HIV infection among high-adherent individuals, regardless of group. A Lancet article that pooled data of transgender women from two clinical trials also showed the efficacy of PrEP in HIV prevention. (DOI: https://doi.org/10.1016	Our conclusion, as it stands, is: Abstract: "PrEP is safe and effective in MSM, serodiscordant couples and PWID. Additional research is needed prior to recommending PrEP in heterosexuals". Discussion: "In conclusion, high-certainty evidence exists that PrEP is safe and, assuming adequate adherence, effectively prevents HIV in MSM and serodiscordant couples. One study found PrEP to be effective in PWID. The uncertainty regarding PrEP effectiveness in heterosexual individuals persists. Clinicians and policy-makers may decide to recommend PrEP to heterosexual individuals on a case-by-case basis, acknowledging adherence-related issues reported in trials." Our conclusion is consistent with WHO guidelines where PrEP is recommended in all populations at "substantial risk". From this review, PrEP has demonstrable efficacy in MSM and serodiscordant couples, and one study has demonstrated efficacy in PWID. In terms of heterosexual transmission, questions remain due to the poor adherence in identified trials and the generalisability of studies conducted in high-prevalence regions. However, clinicians should assess patients on a case-by-case basis, which would include a thorough risk	

	/S2352-3018(15)00206-4)	assessment to determine if the patient is at “substantial risk” and would benefit from PrEP.	
3	PrEP has rapidly evolved during the past years, including the development of formulations and delivery modalities. It looks like the authors focused on oral regimens only. I would make this point more clear by indicating/stating early in the title and abstract.	We have updated the title and abstract to include ‘oral PrEP’, and throughout the main text wherever possible. While the development of new delivery modalities is interesting, only oral PrEP is licensed in the EU/EEA and the focus of this review was to inform the Irish government on the introduction of an oral PrEP programme.	• Title • Abstract (Page 2)
4	Some statements in the Introduction may need to be updated. For example, the latest estimated number of new HIV infections globally is 1.5 million in 2020, which could be found on UNAIDS’s website.	The introduction has been updated with 2020 data relating to global HIV incidence (UNAIDS). Additionally, recent incidence data in Ireland has been added to the discussion section to add local context.	• Introduction, Page 5, 1st paragraph • Discussion, Pages 33,34, under subheading “Research in context and implications for practice”

VERSION 2 – REVIEW

REVIEWER	Larmarange, Joseph Centre Population et Développement, Institut de Recherche pour le Développement, Université Paris Descartes, ERL Inserm U 1244
REVIEW RETURNED	18-Oct-2021

GENERAL COMMENTS	The manuscript has been greatly improved and is now easier to follow and clearer. The authors have taken into account the recommendations of the five reviewers and have provided a detailed point-by-point response letter.

REVIEWER	Dronavalli, Mithilesh University of Western Australia, WACRH
REVIEW RETURNED	06-Oct-2021

GENERAL COMMENTS	I still strongly believe that Bekker 2018 should be excluded from the analysis: The research question may have been mis-specified but in retrospect alternative dosings of PrEP should not be included in the treatment as usual arm.

	Regardless: Bekker2018 is the only study in the heterosexual meta-analysis which uses antiviral medication in the treatment as usual arm. This makes it different to all the other 4 studies that use no treatment in the treatment as usual arm. This makes the meta-analysis highly heterogenous by definition regardless of statistical heterogeneity indicators, eg: I^2 For your research conclusions to reflect your results you need to carry out this meta-analysis for heterosexual people without Bekker 2018. There may be sufficient evidence for PrEP in heterogenous people. Not doing this analysis indicates biased research findings for whatever the underlying reason. Also a not so important point of clarification: Modified intention to treat does not relate to the way patients were recruited. It refers to how the patients were followed up or changed treatments after allocation. Each study should describe how patients were followed up and whether they changed treatments.

REVIEWER	Mostafaei, Shayan Kermanshah University of Medical Sciences, Biostatistics
REVIEW RETURNED	11-Oct-2021

GENERAL COMMENTS	Dear Corresponding Author, Please revise the manuscript according to below comments: - The title is not sharp/clear based on the PICOS. please mention main outcome of the intervention in the title. - In the methods of the abstract, please mention summary of the keywords or search strategy. Also, secondary outcomes should be mentioned in this paragraph. - Please report "[RD]" instead of "[ARD]". - It seems that the term "meta-analysis" could be added in the keywords. - All of applied statistical methods could be mentioned as the sub-section (e.g. statistical analysis) in the methods section. - Please refer to related reference for considering I-squared>0.75 as considerable heterogeneity. I think it will be better heterogeneity among the RCTs has been check by Q (chi-square) test. Then you can decide choosing random effects or fixed effect model for meta-analysis. - The results of sub-group analysis based on the adherence level are not clear. Have the results between groups been same or differ? - Please mention the names of all applied R packages (such as meta r package) in the methods section. - I think the plot of sensitivity plots should be added for clarifying of robustness of the main results. Best regards.

REVIEWER	Li, Chunyan University of North Carolina at Chapel Hill, Health Behavior
REVIEW RETURNED	27-Oct-2021

GENERAL COMMENTS	The authors have addressed the comments from peer reviewers in this revised version. No further comments. Thank you!

VERSION 2 – AUTHOR RESPONSE

Item	Reviewer comment	Response
Reviewer #3:		
1	Bekker2018 is the only study in the heterosexual meta-analysis which uses antiviral medication in the treatment as usual arm. This makes it different to all the other 4 studies that use no treatment in the treatment as usual arm. This makes the meta-analysis highly heterogenous by definition regardless of statistical heterogeneity indicators, eg: I^2 For your research conclusions to reflect your results you need to carry out this meta-analysis for heterosexual people without Bekker 2018. There may be sufficient evidence for PrEP in	We thank the reviewer for the comment and we agree that Bekker 2018 should not be included in the meta-analysis presented for heterosexuals; in fact, Bekker 2018 was not included in the meta-analysis of placebo-controlled trials. In this meta-analysis, four studies were included (Mazzarro 2015, Peterson 2007, Thigpen 2012, VanDamme 2012). The forest plot of these four studies is in the Supplementary Material (Figure S4, Meta-analysis: HIV acquisition in heterosexual participants, PrEP versus placebo) The study by Bekker 2018 was included in the overall review as it was identified by our search. However, as the comparison was different to the placebo-controlled studies, the study results were described narratively. If we had retrieved a number of studies similar to Bekker 2018, it may have been possible to separately estimate the relative effectiveness of PrEP, comparing one dosing schedule with another. We have amended the wording in the text to avoid confusion, including references. The following are the tracked changes: *Of the five studies enrolling heterosexual participants, four were placebo-controlled[7 16 17 19] and one compared different drug schedules.[24] Four studies enrolled only women[7 17 19 24] and one study enrolled both men and women.[16] All studies were conducted in a high HIV prevalence context (countries in Sub-Saharan Africa). A meta-analysis of all the four placebo-controlled studies did not demonstrate a statistically significant reduction in HIV acquisition (RR 0.77, 95% CI: 0.46 to 1.29; I^2 = 66%, Figure S4, Supplementary Material 3.4).[7 16 17 19]*

	heterogenous people. Not doing this analysis indicates biased research findings for whatever the underlying reason.	
	Also a not so important point of clarification: Modified intention to treat does not relate to the way patients were recruited. It refers to how the patients were followed up or changed treatments after allocation. Each study should describe how patients were followed up and whether they changed treatments.	We thank the reviewer for the comment. We agree that modified intention-to-treat does not normally reflect recruitment. However, we were limited by the terminology used in trials, whereby modified intention-to-treat was consistently used to indicate analyses on individuals who were HIV negative at enrolment (and excluding those who were HIV positive prior to receiving study drug). To avoid confusion, we adopted the terminology used in primary studies.
Reviewer #4:		
1	- The title is not sharp/clear based on the PICOS. please mention main outcome of the intervention in the title.	We thank the reviewer for the comment. Our current title is as follows: *Oral Pre-exposure prophylaxis (PrEP) to prevent HIV: a systematic review and meta-analysis of clinical effectiveness, safety, adherence and risk compensation in all populations* In the title we have included the primary outcome of clinical effectiveness of PrEP to prevent HIV. We agree with the reviewer that the following may represent a clearer title: '*a systematic review and meta-analysis of the clinical effectiveness of oral Pre-exposure prophylaxis (PrEP) to prevent HIV'*, however this excludes safety, adherence and risk compensation, which we considered important outcomes to include in the title as they indicate the novelty of the study.
2	- In the methods of the abstract, please mention	We thank the reviewer for the comment. Due to word count limitations, it was difficult to incorporate all feedback regarding the abstract. We have now made amendments to the abstract to summarise key

	summary of the keywords or search strategy. Also, secondary outcomes should be mentioned in this paragraph.	search strategy terms, and we have included secondary outcomes. If the Editorial team wishes to allow a small increase in the number of words allowed (e.g. 310 words), the research team could keep sentences that have been deleted to adhere to word count (these are visible in the tracked changes copy).
3	- Please report "[RD]" instead of "[ARD]"	We thank the reviewer for the comment. This change has been incorporated throughout manuscript.
4	- It seems that the term "meta-analysis" could be added in the keywords.	We thank the reviewer for the comment. 'Meta-analysis' has been added to the keywords.
5	- All of applied statistical methods could be mentioned as the sub-section (e.g. statistical analysis) in the methods section.	We thank the reviewer for the comment. A sub-section for statistical analysis has been added to the methods section.
6	- Please refer to related reference for considering I-squared>0.75 as considerable heterogeneity. I think it will be better heterogeneity among the RCTs has been check by Q (chi-square) test. Then you can decide choosing random effects or fixed effect model for meta-analysis.	We thank the reviewer for the comment. The reference for our assessment of I^2>75% as considerable heterogeneity has been added (Cochrane handbook: https://handbook-5-1.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.html). In line with guidance in the Cochrane Handbook, we have given preference to I^2 as a measure of heterogeneity. Given the limited number of studies included in our review, the Q test has limitations. We would also like to stress that I^2 and Q are measures of statistical heterogeneity, and that clinical heterogeneity is also a very important consideration when pooling studies. While we have given preference to the I^2 test, the chi-square value is reported in the forest plot figures.
7	- The results of sub-group analysis based on the adherence level are not clear. Have the results between groups been same or differ?	We thank the reviewer for the comment. Subgroup analysis based on adherence was performed wherever possible. In MSM studies, PrEP was effective for both high and low adherence subgroups (although the RR was higher in the high adherence group, as expected). In the heterosexual group, three of the four studies had low adherence; the subgroup analysis for this comparison is included in the supplementary material (Figure S5. Meta-analysis: HIV acquisition in heterosexual participants, PrEP versus placebo, studies with low); results were non-significant. Subgroup analyses were not possible for other populations as there were too few studies.

		We have made amendments to the text to clarify these subgroup analyses.
8	- Please mention the names of all applied R packages (such as meta r package) in the methods section.	We thank the reviewer for the comment. We have included reference to the applied R package 'meta package' in the methods section. Reference: Balduzzi S, Rücker G, Schwarzer G (2019), How to perform a meta-analysis with R: a practical tutorial, Evidence-Based Mental Health; 22: 153-160).
9	- I think the plot of sensitivity plots should be added for clarifying of robustness of the main results.	We thank the reviewer for the comment. The plot for publication bias (funnel plot, previously Figure S3.3) has been moved from the supplementary material to the main text (now Figure 2).
Reviewer #1:		
1	The manuscript has been greatly improved and is now easier to follow and clearer. The authors have taken into account the recommendations of the five reviewers and have provided a detailed point-by-point response letter.	We thank the reviewer for their comments.
Reviewer #5:		
1	The authors have addressed the comments from peer reviewers in this revised version. No further comments. Thank you!	We thank the reviewer for their comments.